



A SYNTHETIC POLYNUCLEOTIDE ENCODING HUMAN LACTOFERRIN,  
VECTORS, CELLS AND TRANSGENIC PLANTS CONTAINING IT

DESCRIPTION

Field of the invention

The present invention relates to the field of the vegetable biotechnologies, and in particular to the plants and vegetal cells transformation techniques and systems, to cells and transgenic plants thus generated, and to their use.

State of the art

In recent years, due to the large variety of applications deriving from the utilization of the genetic engineering techniques in the biology of the vegetals, the use of genetically modified plants has gradually increased.

In particular, the development of techniques for the transformation of plants into organisms capable of producing proteins of commercial interest has acquired a remarkable practical importance.

In fact, the generation of recombinant plants containing a heterologous gene of interest, and their use in production processes on an industrial scale, allows to overcome a series of drawbacks characterizing the production systems presently in use, particularly those based on cell cultures.

Indeed, recombinant DNA technology allowed the generation of transgenic cells that are used in the production of heterologous proteins of interest. In particular, animal (and specifically mammal) cell cultures allow the production of proteins of interest, even extremely complex ones, in their native form, but the related process is extremely expensive, as hectoliters of cells are required for the production of an amount of proteins sufficient for a commercial (e.g., pharmacological or alimentary) use (Stowell et al., 1991).

An alternative in this direction is then provided by

the production, carried out in prokaryotic cell cultures, a cheaper process, meeting however a serious obstacle in the inability of those systems to effect the post transcriptional and post-translational modifications required for the expression of complex heterologous proteins, carried out only by the eukaryotic systems (Glick and Pasternak, 1994).

Therefore, a possible solution was pursued in the transformation of complex eukaryotic systems which could ensure at the same time the production of active proteins and the utilization of inexpensive processes. (Watson et al. 1992).

In particular, plants possess the required potential, and their capacity of functioning as bioreactors for the production of complex proteins (plants are highly efficient in this type of processes, ensuring a high degree of expression), in a cost-effective way (plant cultivation is relatively inexpensive) and with a high yield (each single harvest can yield high amounts of proteins) was proved (Fraley and Schell, 1991).

Further, very many plants satisfy the important requirement of the non-allergenicity needed for the production systems of recombinant proteins of pharmacological or alimentary interest. In fact, the organisms used in those productions must belong to the GRAS (Generally Recognized As Safe) organism category, i.e. organisms that having been used by man for a long time, are considered safe for man and for the ecosystem as well. Obviously, plants such as Leguminosae, cereals, tobacco, horticultural plants in general and fruit trees, satisfy this requirement in nature. Among leguminosae, Soya is a basic food in the diet of many populations mostly of the Third World Countries, but recently in European Countries as well. In fact, Soya-derived compounds constitute ingredients that are present in a vast number of food products, such as e.g., lecithin and

seed oil, while the seeds of this Leguminosa yield a flour that is employed in various food like soybean milk containing approximately the 8% thereof.

Therefore plants, while being the raw material from which the product can be extracted according to conventional processes, at the same time constitute an alternative to the conventional production. In fact, they can be used as functional foods, i.e. foods genetically modified so as to be enriched from a nutritional point of view, and in case assuming important properties as a natural drug.

Thus the step of the protein extraction from the plant has been eliminated, and the heterologous protein expression does not prelude to an extraction and purification thereof, steps that account for most of the production cost of a drug, but simply enriches a vegetable nutriment, which thus becomes a *nutriceutical*, i.e. a nutriment having a pharmaceutical value.

This is the reason for the research efforts aimed at generating genetically modified plants, optimized for the above-mentioned applications.

However, to date the evaluation of the heterologous expression capacity typical of various plant species was exclusively performed under laboratory conditions, or anyhow on relatively limited surfaces in hothouse.

For instance, expression in plants of enkephalin and human serum albumin, as well as of mice monoclonal antibodies was studied (Watson et al., 1992). More recently, always referring to proteins relevant from a pharmacological point of view, two additional human proteins of therapeutic importance, i.e. active interleukin-6 and C protein (a serum anticoagulant), were successfully expressed in tobacco. In all these cases the model plant on which the functionality of the prepared gene was tested is tobacco, whereas usually the plant finally selected for production is a leguminosa, whose seed content in storage proteins is high.

These experiments proved that often the expression levels of the heterologous protein in the vegetal bioreactors do not reach high enough levels to meet the commercial demands, and that anyhow it can be improved applying new information on the plant gene control. Specifically, it was demonstrated that the level of produced recombinant protein has to reach more than 1% of the total protein amount in order to become economically significant, a level obviously not achieved by the simple introduction of the heterologous gene, whose expression therefore needs to be maximized.

In relation to the vegetal cell biology, in order to maximize the level of the in plant expression, an action on various levels is necessary: increase of genic transcription, increase of transcript stability and of the translation process yield. Moreover, it is further necessary to fix the inserted gene and to minimize the risk of the occurrence of silencing or of genic inactivation. All these factors are crucial in assessing the in plant expression level of the heterologous gene, that, as afore stated, apart from some exceptions usually settles at rather low levels (Owen and Pen, 1996).

Among all these, the most crucial factor, together with the transcription level depending on the preferential presence of certain codons, would seem to be the stability of the recombinant protein in the heterologous host, as the likely probable cause of its easy elimination. The relative instability might be the consequence, on one hand of the inability of the translated heterologous protein to assume a stable structural conformation, on the other hand of the ultrastructural compartment in which it is directed after the translation, where the presence of proteases and of particular pH values determine its degradation.

Accordingly, it is therefore important both to provide the modification of the heterologous sequences in order to ensure the codon optimization and to carry out a



Careful selection of the targeting sequences, capable of directing the translated protein into preferential ultrastructural compartments, as i.e. the seed storage tissue, capable of ensuring the stability of the product. Concerning the latter issue, the best strategy of action in this sense, is that of providing the maintenance of the endogenous signal sequences of the plant selected for the final production. In fact, the adoption of these sequences prevents the alteration of the cell internal balance, consequence of the unavoidable random accumulation of exogenous proteins that would take place in their absence. In this regard, the fact that the ultrastructural compartments have different characteristics in the cells of the various tissues has to be taken into account.

For instance, the heterologous proteins accumulated together with the storage proteins of the seeds in the transgenic plants, are more stable as compared with those of the vegetative organs (Owen and Pen, 1996). Probably, the reason of such different stability can be found in the different protease activities recorded in the vacuoles of the leaf meristem cells, with respect to those observed in the vacuoles of the seed. It is therefore understandable how research efforts are aimed at the generation of transgenic plants in which the heterologous protein be preferably expressed in the seed (Owen and Pen, 1996).

Furthermore, overall the seed constitutes the vegetal organ most used by man for its ease of conservation and, obviously, for its caloric and protein contribution. The seed consists of the plant embryo enveloped by storage tissues that provide energy and nitrogen during germination. Usually the storage tissue function is mainly effected by the endosperm, but in Leguminosae, as e.g. Soya, the cotyledons develop remarkably, and acquire this function.

The storage function for the nitrogen component is

carried out by particular proteins accumulated in the protein bodies, inside compartments in the endocellular membrane systems. The amount of proteins in the seed varies from about 10-15% in cereals to about 25-35% in Leguminosae, therefore seeds are an important protein source in the human diet (Shewry et al., 1995).

In order to exploit the expression system of the seed storage proteins, first of all it has to be considered that the storage proteins of all plants have some functional and physiological common characteristics: their synthesis is controlled during the seed ripening according to the nutritional needs, and they are stored in protein bodies. In particular, the Leguminosae storage proteins are divided in two classes: globulins and lectins. Globulins are the most widespread storage proteins, present not only in leguminosa but in monocotyledons and in other dicotyledons as well. The globulin class in turn is subdivided into two subclasses: legumins (11S hexameric proteins) and vicilins (7S trimeric proteins). Also  $\beta$ -conglycinine and basic 7S globulin, whose regulation elements were used to perform tissue-specific expression in seed (Shewry et al., 1995) belong to the vicilin subclass, but whereas the former was studied in detail, no detailed information is available on the basic globulin functioning.

$\beta$ -conglycinine is a storage protein of the Soya seed (*Glycine max.*), consisting of three different subunits,  $\alpha$ ,  $\alpha'$ ,  $\beta$  that interact non-covalently to form a trimer complex. The subunits are coded by a multigene family of 15 genes grouped in six nuclear DNA regions, whose expression is strictly regulated so as to be modulated during the plant life (Harada et al. 1989). Control is tissue-specific as well as stage-specific. In fact, the expression of each subunit is activated at high levels at the moment of embryo development, from the heart shaped phase until complete ripening, whereas it is repressed before the dormancy phase.

Moreover, expression occurs exclusively on specific plant zones, like cotyledons, according to an expression pattern that varies in the course of their ripening; at first it is activated on the outer cotyledon area, then a wave distribution from the outside to the inside is observed and lastly it turns out to be uniformly spread over the entire cotyledon tissue. However, during the heart shaped phase (18 days after pollination), the gene encoding the  $\beta$ -conglycinine also expresses itself in the embryonal axis, whereas it does not express in endosperm, tegmen and in already differentiated tissues (Perez-Grau and Goldberg, 1989). The same behavior also occurs in the seed of transgenic plants belonging to other families, as e.g., tobacco, proving the same control mechanism to be functional in solanaceae as well (Naito et al. 1988; Perez-Grau and Goldberg, 1989).

The regulation of subunits  $\alpha/\alpha'$  and  $\beta$  genic expression occur at a transcriptional as well as a post-transcriptional level (Harada et al., 1989).  $\alpha'$  subunit, of 76 Kd, with a 2.5 Kb mRNA, is accumulated more precociously and in a larger amount as compared to  $\beta$  subunit. This behavior is due to the greater strength of the  $\alpha'$  subunit due to the presence of an enhancer region, absent in  $\beta$ , and of a sequence stabilizing the  $\alpha'$  transcript, it also absent in  $\beta$  (probably a 560 pb region in the first exon of the  $\alpha'$  transcript) (Harada et al., 1989). This difference in expression has also been highlighted in transformed tobacco plants (Bray et al., 1987).

In contrast with the  $\alpha'$  subunit, the expression level of the  $\beta$  subunit, is also influenced by abiotic stresses, methionine level in the ground and presence of ABA. The base elements involved in gene control at transcriptional level for the  $\alpha'$  subunit are clustered in the 905 pb region at 5' of the transcription start site, region called URS (Upstream Regulatory Sequence). Inside this area specific sequences functioning as site-specific

enhancers have been detected. Among these, the legumin boxes (5'-CATGCAC.3' and 5'-CATGCAT-3'), elements that are found in many other genes encoding seed-specific proteins, in particular in legumes. A coordinated action of the two sequences determines a 10-fold increase of the seed gene expression level. The regulation by the above elements seems to be of a positive kind (Chamberland et al. 1992), however so far trans-elements specifically interacting in those sites have not yet been found. Site-specific expression also requires the coordinated action of elements operating in cis, not yet characterized, located in the region at 5' of the legumin boxes and at 3' of the promoter (Chamberland et al., 1992). Probably the region at 5' includes a negative control site, specific for a nuclear factor present only in non-seed tissues. This factor would determine gene expression in embryonal tissues only. The importance of CAAT and TATA sequences in the control of site-specificity has also been proved.

Four Soya nuclear factors that interact with specific sequences present in the  $\alpha'$  and  $\beta$  URS of the gene have been identified as well. Two of those embryonal factors, SEF3 and SEF4, bind to sites inside the enhancer region (from -257 to -79). SEF3 binds to the middle of the sequence 5'-AACCCA.....AACCCA-3', present exclusively in the gene encoding the  $\alpha'$  subunit. Accumulation and degradation of this protein (SEF3) is paralleled by accumulation and degradation of  $\alpha'$  mRNA, supporting the hypothesis that SEF3 be involved in the control of  $\alpha'$  expression. As compared to SEF3, SEF4 is present in a lesser amount, has many recognition sites (5'-A/GTTTTTA/G-3') in  $\alpha'$ , but mostly in  $\beta$ . The presence of this factor is correlated to the regulation of  $\beta$  expression (Lessard et al., 1991). Deletion and site-specific mutagenesis experiments have proved that the sole action of these factors does not affect the site-specificity nor the expression level, coordination with



the activity of the other regulatory elements being necessary. However, on the basis of the homology with light-induced proteins, these proteins are supposed to have a regulating role only under certain conditions (Fujiwara and Beachy, 1994).

Embryonal factors with a behavior similar to the SEF3, as verified in gel shift experiments, have been found in tobacco as well (Lessard et al., 1991). This and other data obtained with GUS activity assays under control of Soya  $\alpha'$  promoter prove that the site- and stage-specific control mechanism is conserved (Lessard et al., 1991; Riggs et al., 1989). It has also been hypothesized that tobacco trans factors binding to 35SCaMV promoter may also interact with the legumin boxes (Katagiri et al., 1989). None of the afore mentioned factors appears to be directly responsible of the time regulation, and no NRS-like factor has been found possessing a negative control as in the case of bean (Bustos et al. 1991).

More recent studies, concerning some legumins and vicilins in *Vicia Faba*, contradict some generalizations on the regulation of the storage proteins expression in seeds (Wobus et al., 1995), showing that expression of genes of B4 and LeB4 *Vicia Faba* legumins is not limited to embryonal tissues, nor are they temporally restricted to the cell expansion phase in embryogenesis. Proteins are stored for short periods of time and then degraded in all embryonal tissues, suspensor and endosperm included, within well-defined developmental stages. This is so probably in order to allow an uninterrupted supply to the embryo of compounds having a high C and N content during all the growth and differentiation stages. Therefore, this data allow to hypothesize that the seed proteins expression be also controlled metabolically, and not merely at a developmental stages level. The possible relationship existing between storage proteins accumulation and carbohydrate metabolism (soluble glycid

level) is presently being investigated. Since all classes of seed storage proteins share a similar behavior in the different species, this data require a careful evaluation of the behavior, in terms of expression, also for the Soya proteins  $\beta$ -conglycinine and 7S basic globulin. Data resulting from the study, which the present invention is based on, clearly show the tissutal specificity of the expression of the structural portion of lactoferrin under control of both the promoters. Instead, the activation phase was at the present not investigated in detail as the sole capability of total seed accumulation was of interest. Specifically concerning the 7S basic globulin, it is a storage protein of the Soya seed, with a high methionine and cysteine content. Alike  $\beta$ -conglycinine, also 7S basic globulin (Bg) is stored in seed in large amounts (3% of seed total proteins). It consists of two subunits, one of 27 KDa, the other of 16 KDa encoded by the same mRNA, linked by disulfide bridges. Bg is synthesized as sole precursor polypeptide consisting of a putative peptide signal and of two subunits. This polypeptide is processed to yield the mature dimeric protein. In the genome about four copies of the Bg gene are present (Watanabe and Hirano, 1994).

This protein is mainly located in the seed embryonal tissues and its expression pattern is unusual for a storage protein. In fact, a portion of Bg is accumulated in the intercellular spaces of the cotyledon parenchyma (Nishizawa et al., 1994), whereas at an intracellular level it is stored in protein bodies on the middle lamella of cell wall and in the plasma membrane (Watanabe and Hirano, 1994). This location suggests that the Bg is not a mere storage protein, having other functions as well. More accurate data on Bg location and expression period in Soya are not available. It has never been verified whether the site- and time-specific expression mechanism be preserved in other transformed vegetal species (like tobacco). To this end, reference is made to

general data on storage proteins and to studies on Bg homologues in lupine (conglutyn  $\gamma$ ), with which it has a high sequence homology. This protein is stored only in lupine embryonal tissues (cotyledons and embryonal axis) 40 days after blooming. It has not been detected in other tissues such as leaves and sprouts. In seeds of transgenic tobacco, the conglutyn  $\gamma$  gene is increasingly expressed from the 15th to the 20th day after blooming until the 40th, then begins to decrease (Ilgoutz et al., 1997).

One of the peculiar features of the Bg is that it is secreted in large amount from Soya seeds soaked in water at 40-60°C. It is uncertain whether the secreted proteins are neosynthesized after heat-treatment, or instead are the proteins already present to be secreted. Since a post-heating increase in specific mRNA has been highlighted, it is assumed that the Bg is actually synthesized as a consequence of the thermal shock (Hirano et al., 1989).

Not much is known on the regulation mechanism of the expression of the gene encoding the Bg protein, nevertheless sequences in the promoter region involved in the gene regulation have been identified. Besides the CAAT and TATA box sequences, respectively located at -116 and -25 with respect to the transcription start site, three regulatory elements similar to thermo-specific sequence enhancers present at the non transcribed 5' region of genes in other organisms, have been detected. These heat shock elements (HSE) consist of two 5 pb conserved units: 5'-NGAAN-3' and 5'-NTTCN-3'. In the thermoregulated promoter of the Soya heat-shock protein, the enhancer elements, observed also in Bg as well, cooperate synergistically with three CCAAT box sequences located upstream thus increasing gene expression; these putative sequences are present also in the bg promoter.

Sequences responsible of the site- and time-specificity expression were not identified.

The interest for this protein derives from the fact that it is accumulated in high amounts in the Soya seed (3% of total proteins) and therefore has a strong seed-specific promoter which can ensure a high level of expression of the gene it controls. Moreover, it is known that the regulation mechanism of this protein is different from that of the other storage proteins of Soya seed but the details are not known. However, studies on the promoter and on its site- and time-specific activation mechanism have never been carried out using reporter genes in transgenic plants.

Both Bg and CONG, as storage proteins, are synthesized exclusively in the seed tissue and are stored in large amounts in cells constituting this organ, inside specific compartments. Concerning the post-transcriptional and above all the post translational regulation level, it runs through the mechanisms of intracellular transport and of protein compartmentalization, which are to date to be clarified in many aspects.

In fact, those mechanisms involve all processes influencing the concentration, retention and distribution of the proteins in the endomembrane system (Okita, 1996).

However, general principles of protein targeting do exist, valid for all plant species.

1. Targeting information are contained in the proteins themselves, as discrete signals. Those signals are intercepted by specific recognition signals such as receptors or simple interactions with membrane lipids.

2. Different types of signals do exist (topogenic sequences) each with a specific function. Among them there are signal sequences that start the protein translocation across specific endomembranes and interact with receptors/translocators facilitating the unidirectional transfer. Then there are stop and retention sequences that block the transfer to the membrane or to the inside of the compartment. Selection



sequences target proteins to the various cellular compartments. All those elements can be of a sequential type, i.e. localized in the N-terminus, central or C-terminus portion of the protein, or conformational, i.e. consisting of amino acids which although nonsequential, are yet adjacent in the native protein conformation. Moreover, there may be various signals simultaneously, and they can be modified or activated (e.g. by phosphorylation). After transfer the signal is often deleted using specific cleavage sites for endogenous proteases.

mRNA accumulation in a particular region also influences the protein location. Soya seed storage proteins, globulins as well as lectins, are stored in storage vacuoles. In fact, several types of vacuoles do exist. Some of them, besides having the function of maintaining the turgor pressure and of regulating ion, sugars and amino acid release, also constitute the depository of storage and defense proteins. The specific signal sequence for targeting to the vacuole has not been identified yet, apart from some plant species (Kermode, 1996). Probably, one or more surface regions of the correctly conformed protein are recognized by the selection mechanism. Plant cells possess the unique feature of accumulating storage proteins in the protein bodies, whereas in animals similar inclusion bodies are formed only when an excess of protein synthesis occurs. Therefore, the latter protein bodies consist of unmuddled accumulation of conformationally incorrect or partially processed proteins. The formation and organization process of the protein bodies in plants remains unclear, although it is known that it consists of a series of ordered events (Okita, 1996). Globulins are proteins with an acidic pI, accordingly they are translocated in the ER and in the Golgi complex as soluble proteins. As soon as they reach the vacuoles, due to their low pH or possibly to the processing and assembly, these proteins

precipitate forming particled aggregates that will thereafter originate the protein bodies (Kermode, 1996).

In leguminosae, different storage proteins are accumulated in the same protein bodies with no spatial segregation. In other plant species the protein bodies form in the ER and are then absorbed in vacuoles by autophagocytosis (Kermode, 1996).

This general pattern is well-grounded for  $\beta$ -conglycinine as well, though the specific vacuole targeting sequence have not been identified for this protein. Instead, the  $\beta$ -conglycinine binding with a BIP-homologous protein has been observed. This protein functions as chaperonine and, just like other proteins in different plant species, can have the role of retaining  $\beta$ -conglycinine in the ER until its correct conformation is reached (Galili et al., 1993; Shewry et al., 1995; Kermode, 1996; Fontes et al., 1996). As for 7S basic globulin, available information is scarce. It is known that it is located in protein bodies on the middle lamella of cell wall and in plasma membrane and not in vacuoles (Watanabe, 1994). For this reason, its mechanism of division into compartments is hypothesized to be different from that of  $\beta$ -conglycinine. However, it is known that even wall-located proteins follow the same transport pathway of the vacuole proteins, i.e. are translated in the ER, then transported to Golgi and lastly secreted to the outside or inserted in the membrane by vesicular traffic.

However, storage proteins expression in heterologous hosts shows that the compartmentalization mechanism is universal. In transgenic plants the seed vacuole storage proteins are correctly targeted. Nevertheless, sometimes transport can be inefficient, especially in vegetative organs with respect to seeds. In tobacco, leguminose storage proteins are correctly targeted to vacuoles both in seeds and in leaves, yet in leaves there is a lower accumulation level. This is so because a difference

exists in the transportation efficiency or because of a different processing rate (different proteases or higher instability). Hence, it can be understood how seed-specific in plant production of a heterologous human protein is a complex mechanism, so that the preliminary verifying of the functioning and efficiency of the expression system, as constructed in a model host organism, constitutes an important step.

It has been seen how tobacco is one of the most widely used plants to this end. Its preferential use in assays derives from the fact that it is one of the better known plants, both in a genetic and in a biological and physiological respect. This, together with the ease of effecting the genetic transformation and the shortness of vegetative cycle, made it become an important model for biotechnological experiments, a model whose transformation specific systems and micropropagation conditions are now better known. Additionally, tobacco possesses the further advantage of a near-complete extensibility of the obtained results to several other plant species, consequence of the high conservation rate of genic control mechanisms, that precisely proved to be usually highly conserved in other plant species, and in particular in leguminosae. Therefore it is particularly suitable for the study of the promoters taken out therefrom, and in particular of their capacity of allowing a gene of interest to be expressed in a controlled way in a transformed plant. Genes of interest are usually those encoding proteins suitable in a pharmaceutical or alimentary field. Accordingly, a heterologous gene of interest for this kind of application is that of the human lactoferrin, a protein belonging to the transferrin family, and as such capable to stably and reversibly bind two iron ions.

In fact, by virtue of its biological functions lactoferrin turns out to be interesting from a nutritional as well as from a pharmacological point of

view. It is present in human milk and has a fundamental role in neonatal feeding, as a matter of fact several biological functions have been attributed to this protein, among which a bactericidal and bacteriostatic activity against a wide range of pathogenic microorganisms and the capacity of increasing iron absorption at the intestinal level (Lonnerdal and Iyer, 1995; Hambraeus and Lonnerdal, 1993). Moreover it promotes cellular growth, controls myelopoiesis and is capable of modulating the inflammatory response (Lonnerdal and Iyer, 1995; Oguchi et al. 1995; Penco et al. 1995).

Therefore, at first, attempts to research in milk of other mammals a protein capable of binding iron and possessing the same properties were carried out.

It has been observed that milk of all mammals contains two types of iron-binding proteins, present in different ratios: transferrin, identical to serum transferrin, and lactoferrin. Human milk has a particularly high lactoferrin content, in fact its concentration in colostrum is of 5-10 mg/ml, although it decreases during lactation to about 1 mg/ml in ripe milk (Hambraeus and Lonnerdal, 1993). However, the amount of lactoferrin is much lower in milk of other animal species, like goat, horse, pig and mouse, In cow's milk for instance its concentration is of about 0.1 mg/ml. In some species such as rabbit, rat and dog, lactoferrin is absent, the prevailing iron-binding protein being instead transferrin.

Further, lactoferrin produced by other non-human mammal species, assumes in each of them different structural characteristics, and therefore different properties.

Human lactoferrin (LFU) is a 78 KDa monomer glycoprotein, with a bilobate structure. There is a high degree of homology between the N-terminus domain and the C-terminus one, both at the amino acidic sequence (37%)



and at the tridimensional structure level. The tridimensional structure has been described in detail by X-ray crystallography (Lonnerdal and Iyer, 1995). The gene encoding LFU has been cloned and sequenced. Genic control mechanisms at a transcriptional and translational level and estrogens and iron role in those mechanisms are also known (Liu et al, 1991). The mature protein consists of a 692 aa polypeptidic chain with a 8.8 - 9 pI. It contains 16 disulfide bridges and shows some resistance to proteolysis (Lommerdal, 1995), has three glycanic polyacetylactosaminic chains bound with N-glycosidic bonds to the amino acidic residues Asn233, Asn476 and Asn545 and the molecular weight of the glycosilated protein is 82 KDa.

One of the most important differences existing among LF present in the various animal species is precisely the glycosidic chain composition. In fact, unlike human LF, bovine LF contains  $\alpha$ -1,3 galattosidic residues and glycans of oligomannosidic type; the role of the glycosidic chains has not been defined yet, however it is possible that glycans protect LFU against attacks from proteolytic enzymes. Each of the two LFU domains is capable of binding tightly, yet reversibly to a ferric ion and at the same time to a carbonate or bicarbonate ion molecule (Hambræus and Lonnerdal, 1993). Iron binding sites in human milk lactoferrin are not completely saturated, but only at 6-8% of their capacity (Stowell et al., 1991).

In recent years, a series of studies aimed at understanding the mechanisms of action and the relation between molecular structure and function of this protein have been carried out. The strategy adopted was that of studying LF molecules structurally altered by site-specific mutagenesis. Accordingly, expression systems of LFU recombinants producing in a simple and inexpensive way a protein as identical as possible to the one purified from human milk were carried out.

However, all heterologous hosts used so far for the LFU recombinant are eukaryotes, as, it being a complex glycoprotein, requires a sophisticated processing apparatus.

In 1991 Stowell et al. cloned the LFU gene in cultured neonate hamster renal cells. An inducible  $Zn^{2+}$ -promoter and the secretion signal of a hamster endogenous protein were used to maximize expression. Concentration of LFU recombinant secreted in the culture medium was of about 20 mg/l, sufficient for crystallization and therefore for structural studies. Characterization revealed that it has the same molecular mass of native LFU maintaining intact the iron-binding site. It only differs from human milk LFU in the glycosidic chains and N-terminus sequence, but these do not influence folding. Such an expression system is highly expensive and not suitable for the production of the amounts of proteins required at an industrial level.

Then LFU was expressed in transgenic mice's milk. In this case the entire animal rather than the cultured cells was transformed, using the regulation sequence of the bovine gene  $\alpha$ -S1 caseine. It was shown that LFU mRNA is exclusively expressed in female mammary gland during lactation. In milk the protein reaches a 0.1-36  $\mu$ g/l concentration. Recently this LFU recombinant has been characterized (Nuijens, 1997), and it has been observed that it has the same molecular mass, N-terminus sequence and immunoreactivity of native LFU. It also maintains the capacity of releasing iron at acidic pH and the bond to bacterial lipopolysaccharides. Also in this case, glycosilation, as well as in vivo bactericidal and antiinflammatory action, are different from the ones in native LFU.

Indeed, a highly significant production system of recombinant LFU is that carried out in *Aspergillus awamori* (Ward et al, 1995). This method, which is patented, yields commercial amounts of recombinant LFU (2

g/l). In order to maximize expression LFU is produced as a fusion protein with part of the structural gene, regulation sequence and secretion signal in the culture medium of the glucomylase. The fusion polypeptide is processed to yield mature LFU by an endogenous peptidase. Glucoamylase is an *Aspergillus* protein expressed in high amounts.

An alternative eukaryotic host is the one used by Mitra and coworkers in 1994. They have transformed tobacco cells in suspension. In the transgenic calluses a protein much shorter than the native LFU and therefore found to be unstable is produced in small amounts. Recombinant LFU shows activity against phytopathogenic bacteria, e.g., *Xantomonas campestri*, *Pseudomonas syringae* and others. In the above-mentioned study the obtaining of entire and fertile plants is not reported.

Recently LFU was also produced in culture in insect cells, using Baculovirus as expression system (Salmon et al., 1997). This is a highly powerful expression system, yielding a recombinant protein identical to the native one, apart from the glycosilation level. Nevertheless it maintains the binding with the specific receptors.

All the above reported expression systems allow to obtain an amount of protein sufficing for functional studies, and in some cases (*Aspergillus*) for commercial uses as well. In the latter case however, safe use of the purified protein, e.g. in milk substitutes for neonates, requires an excessive purification in order to ensure the absence of immunogenic or allergenic substances. However, transgenic plants have never been used to this end, nor products directly suitable in human nutrition have been ever yielded, for instance by the recombinant LFU expression in alimentary plants.

#### Summary of the invention

The present invention relates to a general system of tissue-specific, and in particular seed-specific, accumulation of heterologous proteins, designed and

carried out with the object of maximizing the production while preventing the degradation thereof, by using leader sequences and promoters of Bg and  $\beta$  conglycinine genes. To this end, the structural part of the selected gene may encode proteins having an enzymatic activity, used in human therapy or in industrial processes, or proteins with a general (lactoferrin) or specific (antibodies or antigens) pharmacological activity, or antibody proteins for phytopurification or for the elimination of mycotoxins present in foods.

The present invention also relates in particular to a system that, enabling the in plant tissue-specific expression of the human lactoferrin gene, provides an important solution to the problem of the production of this protein. This system in fact determines the production of plants capable of expressing relatively high amounts of this protein that, in the preferred embodiment providing the expression of a synthetic gene designed by the inventors so as to maximize its in plant expression, reaches industrially relevant levels. Moreover, such transgenic plants allow to avoid the costly product purification processes, as they can be used as *nutraceuticals*, and therefore being directly intaken as alimentary products. Accordingly, also the use for the production of protein flours or of protein extracts yielded from tissues, and specifically from seeds of the afore mentioned transgenic plants, for the production of functional foods or special preparations for children is possible.

These plants can anyhow be used also for human lactoferrin purification, by conventional methods based on chromatography techniques.

This expression system further provides the use of new recombinant vectors, constructed by testing the effect of various leader sequences and processing sites of the mature protein, enabling the production of any protein of interest and in particular of lactoferrin or



of fragments derived thereof, with a tissue-specific protein accumulation, in plants belonging to different families among which leguminosae, cereals, solanaceae, fruit-bearing plants and horticultural produce in general. In particular they are structured so as to have the following functionally linked components: (a) a promoter; (b) a signal sequence; (c) a nucleotide sequence optimized for in plant expression and corresponding to the amino acid sequence of the entire human lactoferrin or to fragments thereof; (d) a polyadenylation signal.

In particular the case concerning plasmid is considered, and wherein regulation elements and signal sequences used are those belonging to two genes encoding storage proteins that are very common in Soya seeds, i.e. a  $\beta$ -conglycinine and a 7S basic globulin, isolated and cloned from the Richland soya variety. They can be used to transform plant cells both by the Agrobacterium method and by direct physical methods (Gelvin and Schilperoort, 1995; DuPont Biolistic Manual, DuPont). The vegetal transformed cells are hence selected with the selection agent provided for the purpose and induced to form entire fertile plants capable of forming seeds, in turn capable of expressing the gene for lactoferrin, and accumulating it as storage protein.

This result was obtained by designing and synthesizing an artificial gene encoding the same amino acid sequence of the natural human gene, but having sequence mutations such that codons most frequently used by the human cell are replaced with those most frequently used by the vegetal cell. This result yielded the change of the 31% of the codons in the original gene (see table 1). The remarkable production of human lactoferrin detected in the various plants transformed with the synthetic gene, with yields going from 1% to 1.8% of the seed total storage proteins, but not in plants having the natural gene, proved the in plant functionality thereof.

Concerning all the above disclosed, object of the present invention is a polynucleotide encoding human lactoferrin, characterized in that it has a sequence totally or partially corresponding to the sequence reported as SEQ ID NO:1 and in that said sequence is optimized for in plant expression, and in particular the case wherein said polynucleotide has fused to the 5'-terminus end a sequence selected from the group comprising the sequences shown as SEQ ID NO: 13 and NO: 14.

Object of the present invention is also the human lactoferrin protein, obtained from the expression of the afore mentioned sequences.

A further object of the present invention is a recombinant DNA vector, in particular a plasmid, comprising at least one sequence of a gene of interest, in particular the gene encoding the complete human lactoferrin, specifically a sequence totally or partially corresponding to the SEQ ID NO:1, operatively linked to regulation elements enabling the tissue-specific expression of said gene. A special case is the one where such regulation elements consist of an "expression cassette for plants" allowing tissue-specific expression of said gene.

Further cases of interest are those wherein the "expression cassette" for plants are constituted by the regulation elements of the protein 7S basic globulin, and in particular when among said regulation elements there is the sequence reported as SEQ ID NO:21, or the regulation elements of the  $\beta$ -conglycinine protein, and in particular when among said regulation elements there is the sequence reported as SEQ ID NO:22, the case wherein the sequence of the gene encoding complete or partial human lactoferrin is operatively linked, or even fused, to a "leader sequence", and the case wherein such leader sequence is selected from the group comprising sequences SEQ ID NO: 13 and NO: 14.

Of particular importance is the case wherein such plasmid is selected in the group comprising vectors pBI, pGEM or pUC.

A further object of the present invention is constituted by the transformation process of vegetal cells wherein the transformation is effected with one of the above-mentioned vectors, the transgenic vegetal cells can be obtained through transformation of wild type vegetal cells with at least one of the same vectors, and by the cells that anyhow contain a gene of interest, or portions thereof, and in particular the gene encoding human lactoferrin operatively linked in an "expression cassette", enabling the tissue specific expression of the gene itself, in particular that of the gene encoding 7S basic globulin and that of the gene encoding  $\beta$ -conglycinine.

A further object of the present invention is constituted by cellular aggregates and in particular calluses characterized in that they are obtainable by the aforesaid cells.

A further object of the present invention are also the transgenic plants obtainable from the aforesaid cells with conventional techniques, or anyhow containing a gene of interest and in particular the gene encoding human lactoferrin, specifically the one with a sequence corresponding to SEQ ID NO:1, operatively linked in an expression cassette enabling the tissue specific expression of the gene itself.

Of particular relevance is the case wherein such transgenic plants are selected from the group comprising solanaceae, cereals, leguminosae, horticultural produce and fruit-bearing plants in general, in particular Soya, tobacco and rice, wherein the gene encoding lactoferrin is specifically expressed in the storage tissues or in the fruit. A further object of the present invention is the use of such transgenic plants as *nutriceuticals*.

A further object of the present invention is

constituted by the production processes of functional foods containing proteins produced by the aforesaid transgenic plants, of vegetal milks, starting from natural and/or concentrated proteins deriving from the above-mentioned plants, and anyhow any human lactoferrin production process, characterized in that it utilizes the aforesaid plants.

Lastly, object of the present invention is also the human lactoferrin obtained from the aforesaid transgenic plants.

The invention will be better described with the aid of the annexed figures.

#### Description of the figures

Figure 1 shows the strategy adopted for the assembly of the synthetic gene encoding human lactoferrin.

Figure 2 shows the map of plasmids pGEM-PGLOB(A) and pGEM-PCONG (B) obtained from the cloning of the Soya promoters in plasmid pGEM (Promega), in which the restriction sites used to derive the plasmids are highlighted.

Figure 3 shows agarose gel electrophoresis analysis of digestion of plasmids pGEM-PGLOB with Sal I (lanes 1, 2, 3 and 4) and pGEM-PCONG CON Sph I (lanes 6, 7, 8 and 9), carried out to test clockwise orientation of the insert. All PGLOB samples tested positive, yielding fragments of the expected sizes. PGLOB 1 is sample 1 selected for the subsequent molecular work. In contrast, PCONG samples did not yield the expected pattern, suggesting the possibility of the presence of errors due to the adopted cloning technique (hypothesis later discarded, see figure 6) or due to isolation of a variant of the control region that in the Richland variety differs from the one of the disclosed sequence (Dare variety). Sample 5 is lambda marker HindIII.

Figure 4 shows agarose gel electrophoretic analysis of the restriction pattern of two plasmid pGEM-PCONG clones with enzymes Nde I (lanes 1 and 2), Rsa I (lanes 3



and 4), and SnaB I (lanes 9 and 10), carried out in order to test orientation and identity of the constructs. Cuts with Nde I, Apa I and SnaB I yielded the expected patterns in contrast to the cut performed with Rsa I; this results are justified from the differences found in sequence and reported in figure 6. In lanes 5 and 6 markers of molecular weight  $\lambda$ -DNA Hind III and Marker IV (Boehringer) respectively are present.

Figure 5 shows electrophoretic analysis of the restriction of various clones of plasmid pGEM-PCONG with Rsa I (lanes 1-6) and Hinf I (lanes 9-14) enzymes in order to test identity of constructs. In both cases the obtained pattern do not mirror the expected ones, but are conserved among the different clones, thereby suggesting their being due to differences in the original sequence with respect to the published restriction map and not to errors in the amplification phase with Taq polymerase or in the cloning. Adopted markers are  $\lambda$ -DNA Hind III and Marker IV.

Figure 6 shows the comparison between the published sequence of gene CONG promoter region and the one cloned in plasmid pGEM-PCONG.

Figure 7 shows a schematic view of the two plasmids resulting from the cloning of the native gene LFU into the two vectors pGEM-T and pBI121, carried out in order to obtain plasmids used later on as transformation control.

Figure 8 shows the map of the two plasmids resulting from synthetic LFU gene cloning into vectors pGEM-PGLOB and pGEM-PCONG, i.e. plasmids pGEM-PGLOB-LFU (A) and pGEM-PCONG-LFU (B), respectively.

Figure 9 shows the map of plasmids pBI-PGLOB-LFU (A) and pBI-PCONG-LFU (B) wherein the restriction sites used are highlighted. In particular, box (A) shows the construction of a plasmid containing the synthetic gene represented in the sequences list as SEQ ID NO:1 and cloned in plasmid pBI101 fused to promoter PGLOB and in

Open Reading Frame with the "leader" of 7S basic globulin.

In contrast, box (B) shows the construction of a plasmid containing the synthetic gene reported in the sequence list as SEQ ID NO: 1 and cloned in plasmid pBI101 fused to promoter PCONG and in open reading frame with the  $\beta$ -conglycinine leader.

Figure 10 shows electrophoretic analysis of the restriction of various clones of plasmid pBI-PCONG-LFU with enzymes Xba I, BamH I, Sac I; samples 4 and 5 test positive. In position 6 molecular weight marker Ladder 1Kb is present.

Figure 11 shows electrophoretic analysis of the restriction of two clones of plasmid pBI-PCONG-LFU with enzymes Sal I (lanes 1 and 2) and with Xba I and Sac I (lanes 4 and 5); Both samples tested positive. Samples 3 and 6 represent the positive controls, i.e. pGEM-PCONG-LFU digested with Sal I, Xba I and Sac I respectively.

Figure 12 shows agarose gel electrophoresis analysis of PCR products from genomic DNA extracted from various plants transformed with pBI-PGLOB-LFU, performed using primers PLT48 and PLT49 for the promoter sequence PGLOB. Positive samples 2, 3, 4 and 5 represent the band of the DNA amplified to 1500 base pairs, while samples 6, 7 and 8 represent the negative control of PCR and the positive control (pGEM-PGLOB) respectively. Molecular weight markers Ladder 1Kb are found at 1 and 9.

Figure 13 shows in box (A) the agarose gel with the genomic DNA of tobacco transformed with pBI-PCONG-LFU (lanes 1-5) or with pBI-PGLOB-LFU (lanes 6-9) cut by enzyme BamH I. M is the molecular weight marker Ladder 1 Kb. Sample 10 shows the positive control pGEM-LFU, not shown in the photo for quantitative reasons. In box (B) the hybridization pattern of human LF on the genomic DNA of the same tobacco plants is shown; samples 1, 2 and 3, belonging to plants PCONG 1, PCONG 3, and PCONG 4 respectively are positive, as is the case for samples 5,

7 and 8, belonging to plants PGLOB 10, PGLOB 3 and PGLOB 4 respectively. It is evident that pGEM-LFU, the positive control (lane 10), was only partially digested as also the super-coiled plasmid forms are present.

Figure 14 shows SDS-PAGE electrophoretic analysis of proteins partially purified from seeds of the transgenic plants tested with Southern analysis of the preceding figure (A) and Western analysis of the same proteins, after transfer to a membrane, using polyclonal antibodies specific for the human lactoferrin (B). In particular, in box (A) SDS-PAGE electrophoretic analysis of total cellular proteins (30 DAP) from mature seeds of transgenic tobacco is shown. In position 2, 3, 4 and 5, 6, 7 the same samples tested positive to Southern analysis, extracted by buffer at pH 2.7 and pH 7.6 respectively, are found. Samples 8 and 10 represent the positive control (milk-extracted human lactoferrin, Sigma) and the negative control (non-transformed plant of the same variety), while in position 9 the molecular weight marker Rainbow (Amersham) is found. In box (B) autoradiography of anti-lactoferrin antibody hybridization with the same proteins transferred to DEAE-nitrocellulose membrane is shown; the sample corresponding to plant PGLOB 10, in position 2 and 5, does not yield a positive signal, although according to Southern analysis it is transformed. All other samples are positive.

Figure 15 shows electrophoretic analysis of raw proteins extracted from seeds and leaves of the transgenic tobacco transformed with the plasmids of which at figure 9. In particular, in box (A) protein coloration carried out with Coomassie blue is reported. In box (B) Western Blotting carried out with human LFU-specific antibodies on the proteins of the gel shown in box (A) after transfer to membrane is reported. In particular, in lane 1 plant PGLOB 1, with leaf-extracted proteins is reported, in lane 2 always plant PGLOB 1, with seed-

extracted proteins is reported, in lane 3 plant PGLOB 3, leaf proteins, is reported, in lane 4 plant PGLOB 3, seed proteins, is reported, in lane 5 plant PCONG 105, leaf proteins, is reported, in lane 6 plant PCONG 105, seed proteins is reported, in lane 7 plant PCONG 105, seed proteins treated with N-deglycosilase F, is reported (see text), in lane 8 the molecular weight marker is reported, in lane 9 the human LFU present on the market treated with N-deglycosilase F is reported, in lane 10 LFU present on the market is reported.

Figure 16 shows Western analysis of LFU protein extracted from human milk and of recombinant protein isolated from tobacco seed, before and after N-deglycosilase F enzyme treatment. Analysis was performed with human lactoferrin specific antibodies. In lane 1 LFU extracted and purified by HPLC from seeds of plant PCONG 105 is reported; in lane 2 a protein as in 1 after a 18-hours treatment with N-deglycosilase F is reported; in lane 3 commercial LFU after a 18-hours treatment with N-deglycosilase F is reported; in lane 4 commercial LFU is reported, a diminution is apparent in the molecular weight of the two enzyme-treated samples (2 and 3).

#### Detailed description of the invention

The strategy adopted for the generation of transgenic plants capable of producing human lactoferrin was developed along two directions: on the one hand, comparative analyses on plant expression systems, particularly tobacco and Soya, have been carried out, in order to have a basis for the designing of a sequence encoding human lactoferrin thereof, sequence optimized to maximize its expression in vegetals. Accordingly in the sequence designing the necessity that the required post-translational modifications for the production in the active form could be effected on the translated protein, and that, both for its conformation and due to its subcellular localization, the protein be sufficiently stable to be accumulated in relevant amounts in the



transformed plants, was taken into account. This proved crucial, having ascertained after various attempts carried out in the past years the impossibility of an in plant production of human lactoferrin using constitutive expression systems (e.g. promoter 35S) as well as promoters inducible by leaves cut. Moreover, besides difficulties related to the type of promoter used, the production level and the stability of the protein were tested to be scarce and depending on a warped preferential use of the codons between the human gene and the plants.

Therefore, a plasmid vector system was developed utilizing vectors containing a newly synthesized lactoferrin gene regulated by tissue- and stage-specific promoters capable of yielding a high gene expression and of accumulating the protein in a stable and efficient way inside seed storage organs. Moreover the selection of leader sequences and the design of the fusion point between those and the structural portion of the mature protein yielded a lactoferrin protein that, in quantitative and possibly also in qualitative terms, has the same glycosilation level and the same amino terminal sequence of the native protein, which is important for some of its functional characteristics.

Concerning the synthetic gene design, all the necessary and possible triplets were modified taking into account their preferential use in the two reference plants, tobacco and Soya. In particular, data represented in table 1 were used.

**TABLE 1**

In carrying out such operation, the value G+C and A+T of the two systems (human and vegetal), the non-tandem repeat of some triplets that may cause shifts in reading, etc. were also taken into due account.

Synthetic LFU gene was then obtained using primers reported in the annexed sequence listing from SEQ ID NO:8 to SEQ ID NO:12 and from sequence SEQ ID NO: 15 to SEQ ID NO: 20 and following the assembling strategy reported in figure 1, consisting in repeated PCR cycles, using for each cycle different pairs of synthetic primers allowing the gradual elongation and the forming of the final sequence as designed.

In parallel, also native LFU gene (wild type) encoding human lactoferrin was cloned, always by PCR technique, starting from a cDNA library of human mammalian tissue (Clontech). The gene was recovered in its structural part lacking the signal peptide and the poly-A site and cloned in pGEM-T to form plasmid pGEM-LFU whose map is represented in figure 7. Primers designed for amplification are reported in the annexed sequence listing at sequences SEQ ID NO:2 and SEQ ID NO:3; those added the restriction site BamHI at 5' and the restriction site SacI at 3'. After checking the sequence, which tested identical to the published one, the yielded natural gene was cloned in vector pBI121, on sites BamHI and SacI under control of promoter 35S (see figure 7), this plasmid (pBI-LFU) was then denominated pBI-35S-LFU and used as control in the genetic transformation experiments and in the subsequent molecular and biochemical analysis.

Concerning the preparation of the recombinant vectors containing the elements that allow the tissue-specific expression of the LFU gene in expression cassettes for plants, we proceeded as follows. In order to obtain the seed-specific expression of the protein the promoters and signal sequences of two genes encoding storage proteins that are very abundant in Soya seeds,

i.e. a  $\beta$ -conglycinine (CONG) and a 7S basic globulin (GLOB) were used.

These regulation sequences were isolated and cloned from Soya, Richland variety.

In particular, to clone the two GLOB and CONG sequences PCR technique (PCR = Polimerase Chain Reaction; Innis et al. 1990) was used. In this case genomic DNA extracted from Soya leaves of Richland cultivar was used. Oligonucleotides used for specific amplification are reported in the annexed sequence listing from SEQ ID NO: 4 to SEQ ID NO: 7.

For the GLOB promoter the cloned region includes the entire regulation sequence and the sequence encoding the signal peptide (leader) plus the first codon of the structural sequence, such sequence is indicated with SEQ ID NO: 13 in the annexed sequence listing. For the CONG promoter the cloned region includes the entire regulation sequence and the sequence encoding the signal peptide plus the first codon of the structural sequence, such sequence is indicated with SEQ ID NO: 14 in the annexed sequence listing. For both regulation sequences the most suitable restriction site for insertion proved to be XbaI (TCTAGA), while downstream proved to be BamHI (GGATCC), both absent from the native and synthetic lactoferrin sequence, as highlighted in the following tables 2 and 3.













DNA template was extracted from Glycine max leaves, Richland variety, and amplification products match sizes expected for GLOB (1515 pb) and CONG (1163) promoter on the basis of EMBL sequence data.

Therefore, starting from fragments amplified by ligation in vector pGEM-T, the two vectors pGEM-PGLOB and pGEM-PCONG, whose map is reported in figure 2, were constructed. Yielded plasmids were tested by restriction analysis performed with several enzymes chosen among those cleaving in a limited number and with an overall sequence distribution (see figures 3, 4 and 5) and a clone for each type was selected and sequenced. Sequenced clones showed to be significantly different from the expected sequence. As an example, a comparison between the data bank promoter CONG sequence and the one obtained sequencing clone pGEM-pCONG is reported in figure 6. A 5% difference was detected, therefore the two promoters can be considered as different.

The synthetic gene for human lactoferrin was cloned at first in plasmids pGEM-PGLOB and pGEM-PCONG, cut with enzymes BamHI-SacI, to form plasmids pGEM-PGLOB-LFU and pGEM-PCONG-LFU respectively, whose map is disclosed in figure 8, and then the construct XbaI-SacI transferred in vector pBI101 cut with the same enzymes. In the event of a plant transformation carried out with physical means, as for rice in our case, not reported here, plasmids pGEM-PGLOB-LFU and pGEM-PCONG-LFU, can be directly used after addition of a terminator, in cotransformation with a vector containing the selection marker (e.g., a PUC-type vector containing the gene for hygromycin resistance). Resulting plasmids pBI-PGLOB-LFU and pBI-PCONG-LFU, whose map is reported in figure 9, were used in the genetic transformation of the plants after an accurate control carried out by restriction with various enzymes to assay the correct integration of the DNA construct (figures 10 and 11). Plasmids pBI-35S-LFU, pBI-PGLOB-LFU and pBI-PCONG-LFU were transferred in A.

tumefaciens EHA105 strain cells, made competent by electrophoresis. Strains containing the three plasmids were used to transform about 450 leaf disks (LD) of tobacco, Petit Avana variety. Formation first of shoots and then of roots was induced from calluses formed on leaf disks (LD) in presence of kanamycin. Once rooted, plants were potted and at least 50 kanamycin-resistant plants were analyzed for each construct.

Plants  $T_0$  were tested by PCR technique (figure 12), assaying the presence of the lactoferrin gene inside the genome of the tested plants; plants  $T_1$  were assayed by Southern analysis (figure 13), that compared to PCR technique allows a more accurate testing of the transgene presence in the genome, and with Western analysis (figure 14) allowing detection of genic product and therefore the functionality of inserted gene.

All plants with native LFU gene under control of promoter 35S led to accumulation of a protein, recognized by LFU-specific antibodies, of a molecular weight lower than 50 KDa. This protein was found in small amounts in young leaves, becoming undetectable in the fully developed leaves. Plants transformed with the two constructs pBI-PGLOB-LFU and pBI-PCONG-LFU produce and accumulate exclusively in seed a protein having a molecular weight of 82 KDa corresponding to the glycosilated human protein as shown by electrophoretic analysis of extracted proteins and by the related Western Blotting carried out with LFU-specific antibodies (see figure 15). Presence of recombinant protein exclusively in the seed and not in the leaves was assayed in all the examined transgenic plants (about 50 for the two constructs) with Western techniques.

Recombinant LFU protein isolated from seed and purified with HPLC technique showed to be identical to the native protein concerning its iron binding capacity and its inhibiting effect towards the examined bacterial strains. Treatment with a deglycosilating enzyme confirms

the presence of posttranslational modifications in all alike, at present at least in quantitative terms, to those present in native lactoferrin as highlighted by Western analysis, the results of which are disclosed in figure 16.

It is therefore evident from the above-reported results that using the native gene, described in the literature, for human lactoferrin under control of the traditional promoters used for genetic transformation of plants, human protein lactoferrin cannot be produced in relevant amounts, in a stable form and with the posttranslational modifications typical of this protein.

So far a general description has been given of the present invention. With the aid of the following examples, a more detailed description will now be given of specific embodiments thereof, with the purpose of giving a clearer understanding of objects, features, advantages and methods of application of the invention.

EXAMPLE 1:

Agrobacterium Tumefaciens-mediated tobacco transformation

Day 1

A small amount of *Agrobacterium tumefaciens* of strain EHA 105, taken from a petri plate culture with a sterile loop so as not to exceed in the amount thereby avoiding subsequent problems in controlling bacterial proliferation on plated leaf disks, was inoculated in 2 ml of sterile LB. Then, from a healthy tobacco plant of Petit Avana variety a leaf showing no alteration whatsoever, conversely showing optimal turgor conditions, was taken. The leaf was briefly washed in bidistilled water to remove surface impurities, immersed for 8 min in a 20% sodium hypochlorite and 0.1 % SDS solution and left to dry under a vertical flow hood. From then on all steps were carried out under hood. In particular, the leaf was immersed in 95 % ethanol and shaken in order to completely wet the pages thereof (letting the petiole emerge) for 30 - 40 sec. The leaf was then allowed to dry

out completely.

Disks were obtained from the entire leaf surface with an ethanol-sterilized punch, let fall on plates with MS10 free of antibiotics; in particular, the ratio of 30 disks per plate was not exceeded.

Next, 2 ml LB + (just inoculated) *Agrobacterium* were poured on plate, and the bacterial suspension was evenly spread over the entire plate with a gentle rotatory movement, in order to obtain an homogeneous bacterial distribution among the disks. LB in excess was carefully aspirated with a pipette. In the course of those steps at all times a parallel negative control was provided by means of a plate to which nothing, or only LB was added.

Then plates were incubated at 28°C for 24-48 hours in constant lighting conditions, and bacterial growth was indicated by the appearance of a thin opaque layer spreading over the entire plate.

#### Day 2

Leaf disks (=LD) were carefully transferred on a plate with MS10 + 500 mg/l cefotaxime, and incubated at 28°C for 6 days in constant lighting conditions. This step determines the *Agrobacterium* inactivation.

#### Day 8

LD were then carefully transferred on a plate with MS10 + 500 mg/l cefotaxime and 200 mg/l Kanamycin, and incubated at 28°C for 14 days in constant lighting conditions. This step determined a selection of the transformed plants: in fact, gene of kanamycin resistance was carried by the plasmid inserted in *Agrobacterium*.

#### Day 22

LD, that in the meantime had grown developing a callus, were carefully transferred on a plate with MS10 + 500 mg/l cefotaxime, 200 mg/l Kanamycin and 500 mg/l carbenicillin, and incubated for 6 days. This step determines elimination of the *agrobacteria* possibly survived to the previous antibiotic treatments (a very frequent occurrence).



Day 28

LD were transferred again on MS10 + 500 mg/l cefotaxime and 200 mg/l Kanamycin, and incubated until shooting. When shoots showed at least two leaves, they were separated from the callous mass and transferred on the radication medium: MS0 + 500 mg/l cefotaxime and 200 mg/l Kanamycin.

At the appearance of roots, seedlings were extracted from the plate, freed from agar residues, gently washed in running water and planted out in loam and sand (2:1) inside small plastic pots. Soil was previously saturated with water, then pots were covered with transparent plastic lids to preserve high humidity conditions, and placed in a growth chamber at room temperature, with a daily 16-hour lighting period.

EXAMPLE 2:

Purification of lactoferrin protein from different tissues of the plant and assessment of molecular weight.

Extraction of all the proteins of tobacco seed was performed grinding the seeds in liquid nitrogen in presence of an extraction buffer (0.5 M saccharose, 0.1% ascorbic acid, 0.1% Cys-HCl, 0.01 M Tris-HCl, 0.05M EDTA pH 8).

Then the solution was centrifuged for 30 minutes at 14.000 rpm at 4°C and the supernatant was kept with the soluble proteins.

Then the solution was filtered with filters of 0.2 µm porosity, and the lactoferrin partially purified by removing proteins of a molecular weight lower than 30 KDa by centrifugation in Centricon 30 column (Amicon).

The lactoferrin was further purified by HPLC chromatography on Resource Q column (Pharmacia) at a weak cationic exchange, with elution in phosphate buffer pH 7 and NaCl gradient 20-100%. The peak corresponding to lactoferrin eluted at 0.7 M NaCl.

The fractions of the elution range were reunited and filtered in Centricon 30 to remove salt.

For the lactoferrin extraction from tobacco leaves, up to the centrifugation step we proceeded as in the case of extraction from seed, then the supernatant was additioned with 60%  $(\text{NH}_4)_2\text{SO}_4$  and left shaking in ice for 60 min.

Then the solution was centrifuged at 14.000 rpm for 15 minutes at 4°C, the pellet recovered and then suspended again in phosphate buffer pH 6.8.

For the assessment of molecular weight in SDS-PAGE, the colorant (SDS loading buffer) was additioned to the lactoferrin sample (20 $\mu$ l) and the samples were loaded onto 8% polyacrilamide minigels. Running conditions were: initially 10mA, and 20 mA for the entire run, in Tris-glycine 1x buffer. Then the gel was stained by Silver staining technique and the molecular weight assessed referring to molecular weight standards.

#### EXAMPLE 3:

Western analysis of the lactoferrin protein produced in plant and deglycosilation thereof.

Lactoferrin purified from seed according to example 2, after electrophoretic separation on acrylamide gel was transferred by electroblotting (buffer 25mM Tris, 192 mM glycine, 20% methanol, 45 V at 4°C) to a nitrocellulose membrane (BA85 Schleicher and Schull).

The membrane with the immobilized protein was shaken for 60 min in TBS-T 5% Skin milk solution and then, after some washings, with the same solution containing the primary antibody in a 1:2500 ratio.

After reaction with primary antibody the membrane was washed and placed in contact with the secondary antibody (Anti-Rabbit peroxidase conjugate), always in TBS-T Skin milk solution, in a 1:12.000 ratio.

After reaction with secondary antibody the membrane was washed several times and placed in contact with Amersham's chemiluminescence kit ECL.

The membrane was then exposed in contact with a photoplate (Hyperfilm MP, Amersham) in darkroom for

variable lengths of time.

Deglycosilation with N-glycosidase F enzyme (Boehringer Man.) was carried out using 10  $\mu$ l in volume of glycopeptide (10  $\mu$ g) denatured in 0.1% SDS brought to boiling point for 2 min.

To this solution 90  $\mu$ l of buffer (20 mM phosphate buffer pH 7.2, 50 mM EDTA pH 8, 10 mM sodium azide, 0.5% NP40, 1%  $\beta$ -mercaptoethanol) were additioned and it was brought to boiling point again for 2 min, then cooled at 37°C.

To the resulting 100  $\mu$ l 1 U of N glycosidase F was additioned and let incubate at 37°C for 18 hours. Then the reaction product was analyzed on SDS-PAGE gel and the lactoferrin protein detected by Western technique.

#### GLOSSARY

The term "recombinant polynucleotide", as it is used here to characterize a polynucleotide useful in the production of lactoferrin, relates to a polynucleotide of genomic origin, cDNA, semi-synthetic or synthetic, that, by virtue of its origin or manipulation: 1) is not associated to a portion or to the totality of the polynucleotide to which it is associated in nature, and/or 2) is linked to a polynucleotide differing from that to which it is associated in nature, or that 3) does not exist in nature.

The term "polynucleotide", as it is used here, relates to a polymeric form of nucleotides of any length, ribonucleotides as well as deoxyribonucleotides. This term exclusively refers to the molecule primary structure. Hence, the term includes single and double stranded DNA as well as single and double stranded RNA. It also includes modified forms of the polynucleotide, e.g. by methylation, phosphorilation or "capping", and non modified forms.

An «expression cassette for plants» relates to a recombinant polynucleotidic sequence obtained by linking together operatively various elements constituted by the polynucleotidic sequences that determine the in plant

expression of a character and that are easily transferable as discrete constructs, from a vector to another by enzymatic restriction.

A "vector" is a replicon to which another polynucleotidic fragment is added, in order to effect the replication and/or expression of the fragment itself.

A "replicon" is any genetic element, for instance a plasmid, a chromosome, a virus, that behaves as an autonomous polynucleotidic replication unit inside a cell; therefore it can replicate autonomously.

"regulation sequence" refers to polynucleotidic sequences that are needed to effect the expression and/or the secretion of coding sequences to which they are bound. The nature of these regulation sequences differs depending on the host; in prokaryotes those regulation sequences usually include promoter, binding site of ribosomes and terminators; in eukaryotes these regulation sequences usually include promoters, terminators and, in some cases, *enhancers*. In addition, in prokaryotes as well as in eukaryotes, leader sequences control the host cell secretion of the expressed polypeptide. The term "regulation sequences" includes, at least, all components whose presence is required for expression, and may also include additional components whose presence is advantageous, for instance leader sequences.

A «leader» sequence is a polynucleotidic fragment, usually short, encoding a transport signal of the protein fused thereto and leading the protein transfer into specific cellular compartments. If the transfer takes place through the endoplasmic reticulum the protein undergoes specific posttranscriptional modifications.

"Operatively linked" relates to a juxtaposition wherein the above described components are in a relation enabling them to function in the expected way. A regulation sequence «operatively linked» to a coding sequence is linked in such a way that the coding sequence expression takes place in conditions that are compatible

with the regulation sequences.

AN *open reading frame*, *ORF* is a polynucleotidic sequence region encoding a polypeptide; this region can represent a portion of coding sequence or a complete coding sequence.

A "coding sequence" is a polynucleotidic sequence that is transcribed in the mRNA and/or translated in the polypeptide when placed under control of appropriate regulation sequences. The ends of the coding sequence are determined by a translation start codon at 5' and by a translation stop codon at 3'. A coding sequence can include, without being limited to, mRNA, cDNA, and recombinant polynucleotidic sequences.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures" and other terms indicating microorganisms or cell lines of superior eukaryotes, cultivated as unicellular entities, are used here in an interchangeable way. They relate to cells that can be, or have been, used as hosts for recombinant vectors or other transfer polynucleotides, including the progeny of the cell that was originally transformed. It is implicit that, due to random or deliberate mutations, the progeny of a single parental cell need not necessarily be identical to the parental cell from a morphological and a genetic point of view. Progenies of the parental cell that are sufficiently similar to the ancestor cell and can be characterized for their salient capacity, as e.g., the presence of a nucleotidic sequence encoding the peptide of interest, are included in the progeny understood according to this definition and fall within the same terms.

For «cell aggregation» a group of cells that are not structured in an organized tissue, but result from an undifferentiated proliferation of cells maintained in particular conditions of hormonal concentration.

"Transformation", as it is used here, refers to the insertion of a exogenous polynucleotide in a host cell, regardless of the method used for the insertion itself,



e.g. direct acquisition, Agrobacterium infection, sexual reproduction. The exogenous polynucleotide can be maintained as a non integrated vector, for example a plasmid or, alternatively, it can integrate in the host genome.

As it is used here, the term "polypeptide" relates to the amino acidic product of a sequence encoded inside a genome and does not relate to the specific length of the product: accordingly, peptides, oligopeptides and proteins are included in the definition «polypeptide» This term does not relate to the post-expressional modifications of the peptide, as e.g. glycosilation, acetylation, phosphorilation, sialilation and the like.

A "wild type polypeptide", has an amino acidic sequence identical to the one encoded in the genome of the organism source of the coding sequence.

"Native lactoferrin " and analogous terms relate to the lactoferrin isolated from the source in which it is usually produced in nature by a genome existing in nature.

A "non-native polypeptide " refers to a polypeptide that is produced in a host differing from the one wherein it is produced in nature.

# REFERENCES

- Anderson B.F., Baker H.M., Norris G.E., Rice D.W., Baker E.N. 1989. J. Mol. Biol. 209:711-734.
- Baeuerle P.A. 1995. Nature 373:661-662.
- Bezault J.A., Bhimani R., Wiprovnick J., Furmanski P. 1994. Cancer Res. Baltimore 54:2310-2312.
- Boesman Finkelstein M., Sciortino C.V., Finkelstein R.A., Spik G., Montreuil J., Chrichton R.R., Mazurier J. 1985. Proc. 7th Int. Conf. Prot. Iron Metab. 251-260.
- Bray E., Naito S., Pan N., Anderson E. Dubé P.H., Beachy R.N. 1987. Planta 172:364-370.
- Bustos M.M., Begum D., Kalkan F.A. Battraw M.J., Hall T.C. 1991. EMBO J. 10:1469-1479.
- Chamberland S., Daigle N., Bernier F. 1992. Plant Mol. Biol. 19:937-949.
- Fontes E.P.B., Silva C.J., Carolino S.M.B., Figueiredo J.E.F., Batista D.P.O. 1996. Braz. J. Genetics 19:305-312.
- Fraley R., Schell J. (eds.) 1991. Curr. Opinion Biotechnol, 2:145-210.
- Fujiwara T., Beachy R.N. 1994. Plant Mol. Biol. 24:261-272.
- Galili G., Altschuler Y., Levanony H. 1993. Trends in Cell Biol. 3:437-442.
- Gelvin S.B., Schilperoort R.A. 1995. Plant Molecular Biology Manual. Kluwer Acad. Pub., London.
- Glick B.R., Pasternak J.J. 1994. Molecular Biotechnology. ASM Press, Washington, p. 113.
- Grover M., Giouzeppos O., Schnagl R.D., May J.T. 1997. Acta Paediatr. 86:315-316.
- Hambraeus L., Lonnerdal B. 1993. Indigenous Antimicrobial Agents in Milk-Recent Developments. Pp.97-107.
- Harada J.J., Barker S.J., Goldberg R.B. The Plant Cell, 1:415-425.
- He J., Fumanski P. 1995. Nature 373:721-724.

- Hirano H., Kagawa H., Okubo K. 1989. 31:731-735.
- Ilgoutz S.C., Knittel N., Min Lin J., Sterle S., Gayler K.R. 1997. Plant Mol. Biol. 34:613-627.
- Innis M.A., Gelfand D.H., Sninsky J.J., White T.J. 1990. PCR Protocols: a guide to methods and applications. Academic Press, New York.
- Kagawa H., Yamauchi F., Hirano H. 1987. FEBS Lett. 226:146-149.
- Katagiri F., Lam E., Chua N.H. 1989. Nature 340:727-730.
- Kermode A.R. 1996. Critical Rev. In Plant Sciences, 15:285-423.
- Komatsu S., Hirano H. 1991. FEBS Lett. 294:210-212.
- Lessard P.A., Allen R.D., Bernier F., Crispino J.D., Fujiwara T., Beachy R.N. 1991. Plant Mol. Biol. 16:397-413.
- Lonnerdal B., Iyer S. 1995. Ann. Rev. Nutr. 15:93-110.
- Mann D.M., Romm E., Migliorini M. 1995. J. Biol. Chem. 269:23661-23667.
- Marchetti M., Longhi C., Conte M.P., Pisani S., Valenti P., Seganti L. 1996. Antiviral Res. 29:221-231.
- Miehlke S., Reddy R., Osato M.S., Ward P.P., Conneely O.M., Graham D.Y. 1996. J. Clin. Microbiol. 34:2593-2594.
- Mitra A., Zhang Z. 1994. Plant Physiol. 106:977-981.
- Naito S., Dubé P.H., Beachy R.N. 1988. Plant Mol. Biol. 11:109-123.
- Nakajima M., Shinoda I., Samejima Y., Miyauchi H., Fukuwatari Y., Hayasawa H. 1997. J. Cell Physiol. 170:101-105.
- Nishizawa N.K., Mori S., Watanabe Y., Hirano H. 1994. Plant Cell Physiol. 35:1079-1085.
- Nuijens J.H., van Berkel P.H., Geerts M.E., Harteveldt P.P., de Boer H.A., van Veen H.A. 1997. J.

Biol. Chem. 272:8802-8807.

- Oguchi S., Waker W.A., Sanderson I.R. 1995. Biol. Of the Neonate, 67:330-339.
- Okita T.W. 1996. Ann. Rev. Plant Physiol. Plant Mol. Biol. 47:327-350.
- Owen M.R.L., Pen J. 1996. Transgenic plants: A production system for industrial and pharmaceutical proteins. J. Wiley e Sons, New York.
- Paule-Eugene N., Dugas B., Kolb J.P., Damais C., Braquet P., Paubert-Braquet M., Rialland J.P. 1993. Comp. Rendus Acad. Sci., Series 3, Sci. De la Vie, 316:113-119.
- Penco S., Pastorino S., Bianchi-Scarra G., Garre C. 1995. J. Biol. Chem. 270:12263-12268.
- Perez-Grau L., Goldberg R.B. 1989. The Plant Cell, 1:1095-1109.
- Petschow B.W., Talbott R.D. 1991. Pediatric Res. 29:208-213.
- Riggs C.D., Voelker T.A., Chrispeels M.J. 1989. Plant Cell, 1:609-621.
- Salmon V., Legrand D., Georges B., Slomianny M.C., Coddeville B., Spik G. 1997. Protein Expr. Purif. 9:203-210.
- Samaranayake Y.H. et al. 1997. APIMIS, 105:875-883.
- Shewry P.R., Napier J.A., Tatham A.S. 1995. The Plant Cell, 7:945-956.
- Soukka T., Lumikari M., Tenovuo J. 1991. Microbial Ecol. In Health and Dis. 4:259-264.
- Steiner T.S. et al. 1997. Clin. Diagn. Lab. Imm. 4:719-722.
- Stowell K.M., Rado T.A., Funk W.D., Tweedie J.W. 1991. Biochem. J. 276:349-355.
- Superti F., Ammendolia M.G., Valenti P., Seganti L. 1997. Med. Microb. Imm. 186:83-91.
- Swart P.J., Kuipers M.E., Smit C., Pawels R., De Bethune M.P., De Clercq E., Meijer D.K.F., Huisman J.G. 1996. AIDS Res, and Human Retroviruses, 12:769-775.

- Teraguchi S., Ozawa K., Yasuda S., Shin K., Fukuwatari Y., Shimamura S. 1994. Biosc. Biotech. And Biochem. 58:482-487.
- Teraguchi S., Shin K., Ozawa K., Nakamura S., Fukuwatari Y., Tsuyuki S., Namihara H., Shimamura S. 1995. Appl. Env. Microb. 61:501-506.
- Tomita M. 1993. Ind. Antimicrob. Agents of Milk-Recent Develop. pp.7-12.
- Van Berkel P.H.C., Geerts M.E.J., van Veen H.A., Mericskay M., De Boer H.A., Nuijens J.H. 1997. Biochem. J. 328:145-151.
- Ward P., Lo J.Y., Duke M., May G.S., Headon D.R., Conneely O.M. 1992. BioTechn. 10:784-789.
- Ward P., Piddington C.S., Cunningham G.A., Zhou X., Wyatt R.D., Conneely O.M. 1995. BioTechn. 13:498-503.
- Ward P., Zhou X., Conneely O.M. 1996. J. Biol. Chem. 271:12790-12794.
- Watanabe Y., Barbashov S.F., Komatsu S., Hemmings A.M., Miyagi M., Tsunasawa S., Hirano H. 1994. Eur. J. Biochem. 224:167-172.
- Watanabe Y., Hirano H. 1994. Plant Physiol. 105:1019-1020.
- Watson J.D., Gilman M., Witkowski J., Zoller M. 1992. Recombinant DNA. W.H. Freeman and Co. New York, p. 213.
- Wobus U., Borisjuk L., Panitz R., Manteuffel R., Baumlein H., Wohlfahrt T., Heim U., Weber H., Misera S., Weschke W. 1995. J. Plant Physiol. 145:592-599.